

Univerzita Karlova, Přírodovědecká fakulta
Charles University, Faculty of Science

Doktorský studijní program: Parazitologie

Doctoral study program: Parasitology



RNDr. Štěpánka Hrdá

Summary of the doctoral thesis

Evoluce jaderných a plastidových genomů u euglenidů

Evolution of nuclear and plastid genomes in euglenids

Školitel / Supervisor: doc. Vladimír Hampl, PhD.

Praha 2020

Abstract

Algae form a diverse group of simple photosynthetic eukaryotes of polyphyletic origin. Algae with a primary plastid (Archaeplastida) acquired it by ingesting cyanobacterium, a prokaryote; algae with a complex plastid acquired their plastid by ingesting another eukaryote with a primary or already complex plastid. Algae with a complex plastid are chimeras containing genes derived from the host genome, as well as genes derived from the genome of the endosymbiont, and also genetic material derived from genomes of their previous stable or transient endosymbionts. One of the groups with plastid derived from green algae are euglenophytes. This thesis deals with the genomes of three organisms that represent individual actors in the endosymbiotic process in euglenophytes. These are a heterotrophic host from the class Euglenida, a phototrophic endosymbiont from the class of green algae Prasinophyceae and the resulting phototrophic euglenid from the group Euglenophyceae. Knowledge of their genomes should illuminate the course of endosymbiotic gene transfer (EGT) in the formation of algae with a complex plastid.

We annotated the plastid genome of a phototrophic euglenid *Eutreptiella gymnastica* and published it as the third plastome of Euglenophytes after the iconic and economically important *Euglena gracilis* and its secondarily non-photosynthetic relative *E. longa*. We have shown that the content of genes in the plastid genomes of Euglenophytes is almost the same and that the reduction of plastid gene content by about 20 % has already occurred in the common ancestor of all euglenophytes. We further found that the plastome of the basal euglenophyte *Etl. gymnastica* has only few introns and therefore the extremely high number of introns in *E. gracilis* plastome is not a common feature of the whole group but a species-specific feature. Last but not least, we confirmed that the green alga *Pyramimonas parkeae* is related to the ancestor of euglenophytes' endosymbiont. The plastid genome of this alga was already known, and we complemented it by the mitochondrial genome with relatively high gene content. None of these genes is probably present in current euglenophytes' genomes, and the entire mitochondrial gene pool of the endosymbiont has been lost during endosymbiosis. The last organism examined in this study is the osmotrophic euglenid *Rhabdomonas costata* whose partial nuclear genome and almost complete transcriptome were compared with the currently available genome, transcriptome and mitochondrial proteome of *E. gracilis*. We found unconventional introns typical for euglenids in *R. costata* nuclear genome. Its transcriptome contains a very low portion of genes related to the genes of photosynthetic organisms and lacks some genes obtained by *E. gracilis* via EGT which disproves the hypothesis of early plastid acquisition in Euglenozoa.

These individual findings undoubtedly contributed to the understanding of the evolution of Euglenida and the collected data will be used in further studies.

1 Introduction

Euglenida is a relatively large group of marine and freshwater protists that involves species with four different types of feeding strategies: bacterivores, eukaryovores, osmotrophs and photoautotrophs (Leander et al., 2001). Phylogenetic studies demonstrate that only photoautotrophs are monophyletic and evolved from eukaryovorous ancestor. Bacteriovores and eukaryovores are polyphyletic. Osmotrophs consist of a monophyletic group of primary osmotrophs (descendants of other eukaryovorous ancestor than photoautotrophs) and a polyphyletic group of secondary osmotrophs that contains the species that lost the photosynthesis (Lax et al., 2019; Lax & Simpson, 2013; Leander et al., 2017). The phototrophic euglenids and their secondary heterotrophic descendants are classified as class Euglenophyceae. The fact that plastids are present in a single clade of euglenids favors hypothesis that the ancestor of this clade acquired the plastid by engulfing a green alga.

Analyses of several plastid genomes has pointed to *Pyramimonas* (Pyramimonadales, Prasinophyceae) as the closest extant relative of the euglenid plastid (Turmel et al., 2009). An alternative but currently less accepted plastid-early hypothesis postulates that the euglenid plastid was acquired early in the evolution of euglenids, or even in the common ancestor of euglenids and kinetoplastids (e.g. *Trypanosoma*), their closest sister group (Hannaert et al., 2003).

Thirty euglenophyte plastid genomes (Bennett et al., 2012, 2014; Dabbagh & Preisfeld, 2018; Gockel et al., 2000; Hallick et al., 1993; Hrdá et al., 2012; Karnkowska et al., 2018; Pombert et al., 2012; Wiegert et al., 2012) and only one mitochondrial genome of *E. gracilis* (Dobáková et al., 2015) have been published to this date. Only partial genomic sequence of *E. gracilis* and transcriptomes of several phototrophic and secondary osmotrophic species are available (Ebenezer et al., 2019; O'Neill et al., 2015; Yoshida et al., 2016; Záhonová et al., 2018), while primarily heterotrophic euglenids are undersampled in this respect. Recently, proteomes of *E. gracilis* plastid (Novák Vanclová et al. 2020) and mitochondrion (Hammond et al. 2020) have been revealed by mass spectrometry proteomics. In this thesis the several genomic studies are presented to complete the picture of evolution of nuclear and plastid genomes in euglenids.

2 Aims

1. To characterize plastid genome of *Eutreptiella gymnastica*, euglenophyte distantly related to the model species *Euglena gracilis*.
2. To characterize mitochondrial genome of *Pyramimonas parkeae*, living relative of euglenophyte plastid ancestor.
3. Based on genomic and transcriptomic data, to characterize selected features of the heterotrophic euglenid *Rhabdomonas costata* and compare them with the features of *Euglena gracilis*.

3 Materials and methods

Materials and methods are described in detail in the respective publications. Briefly, DNA and RNA sequencing were used to obtain extensive sequence datasets which were subsequently examined using bioinformatic tools for organelle genome annotations, intron searching, homology detection, metabolic reconstructions, and determination of phylogenetic affiliation.

4 Results and discussion

We presented the plastid genome of the euglenid *Eutreptiella gymnastica*. Comparison of this genome and in 2012 the only other available plastid genome of photosynthetic euglenid, *Euglena gracilis* (Hallick et al., 1993), revealed that they contain a virtually identical set of protein coding genes. Today, after annotation of 30 euglenophytes plastid genomes it is known that the highly conserved core content of euglenophyte chloroplast DNAs consists of 89 genes (Bennett et al., 2012, 2014; Dabbagh & Preisfeld, 2018; Karnkowska et al., 2018; Pombert et al., 2012; Wiegert et al., 2012). The plastid genome of *Pyramimonas parkeae*, the closest extant algal relative of the euglenid plastid, contains 110 genes (Turmel et al., 2009). The gene content was reduced during the secondary endosymbiosis. Searching within the transcriptomes of *Euglena* and *Eutreptiella* showed that 6 of the missing genes were transferred to the nucleus of the euglenid host while 18 have been probably lost completely. *Euglena* and *Eutreptiella* represent the deepest bifurcation in the photosynthetic clade, and therefore all these gene transfers and losses must have happened in the last common ancestor of all known photosynthetic euglenids. The conservation of gene content in the two lineages

of euglenids is in contrast to the variability of gene order and intron numbers, which diversified dramatically. Our results show that the early secondary plastid of euglenids was much more susceptible to gene losses and endosymbiotic gene transfers than the established plastid, which is surprisingly resistant to changes in gene content.

We reported the mitochondrial genome of prasinophyte alga *Pyramimonas parkeae*. Prasinophytes are a paraphyletic assemblage of nine heterogeneous lineages in the Chlorophyta clade of Archaeplastida (Leliaert et al., 2012). The characterized mitochondrial genome is very compact and the gene content is slightly smaller than the gene-richest prasinophyte mitochondrial genomes. The single identified intron is located in the cytochrome c oxidase subunit 1 gene (*cox1*) and the mature mRNA is formed by trans-splicing. The phylogenetic analyses based on all three genomic elements support the sister position of prasinophyte clades Pyramimonadales and Mamiellales. We have searched for the homologues of *P. parkeae* mitochondrial genes in transcriptomes of *Etl. gymnastica* and *E. gracilis* in order to identify potential endosymbiotic gene transfers from the mitochondrial genome of the plastid donor into the Euglenophyceae host lineage. No close homologues were found, so we have no evidence of such a transfer.

We present genomic and transcriptomic drafts of a primary osmotroph *Rhabdomonas costata*. The current genomic assembly length 100 Mbp is 14× smaller than in *E. gracilis*, the difference partly caused by incompleteness of the data (Ebenezer et al., 2019). The genome is too fragmented for comprehensive gene prediction, nevertheless, the comparison of transcriptomic and genomic data allowed us to estimate features of its introns including several candidates for non-conventional intron types. Transcriptome dataset (39,585) was sufficient for functional annotation. Only 26,052 predicted proteins have any homologue in NCBI and 13,130 were annotated by KEGG. Annotation of the mitochondrial core metabolism provides the first data on the facultatively anaerobic mitochondrion of *R. costata*, which in many aspects resembles the mitochondrion of *Euglena gracilis* (Ebenezer et al., 2019; Hammond et al., 2020). *R. costata* synthesises heme by a mitochondrial-cytoplasmatic C4 pathway with enzymes orthologous to those from *E. gracilis* (Kořený & Oborník, 2011). This, together with a low (overall) percentage of green algae-affiliated genes supports the ancestrally osmotrophic status of this species and disproves the hypothesis of early plastid acquisition in Euglenozoa.

5 Conclusions

We characterized the plastid genome of *Eutreptiella gymnastica* and compared it with that of *Euglena gracilis*. We showed that plastid genome content is similar in Euglenophytes and was established soon after plastid acquisition. We characterized the mitochondrial genome of *Pyramimonas parkeae* and described some of its interesting features. Based on genomic and transcriptomic data we characterized the partial genome and almost complete transcriptome of the heterotrophic euglenid *Rhabdomonas costata* and we predicted *in silico* mitochondrial proteome, tetrapyrrole synthesis pathway and several conventional and non-conventional introns.

6 References

- Bennett, M. S., Wiegert, K. E., & Triemer, R. E. (2012). Comparative chloroplast genomics between *Euglena viridis* and *Euglena gracilis* (Euglenophyta). *Phycologia*, *51*(6), 711–718. <https://doi.org/10.2216/12-017.1>
- Bennett, M. S., Wiegert, K. E., & Triemer, R. E. (2014). Characterization of *Euglenaformis* gen. nov. and the chloroplast genome of *Euglenaformis* [*Euglena*] *proxima* (Euglenophyta). *Phycologia*, *53*(1), 66–73. <https://doi.org/10.2216/13-198.1>
- Dabbagh, N., & Preisfeld, A. (2018). Intrageneric Variability Between the Chloroplast Genomes of *Trachelomonas grandis* and *Trachelomonas volvocina* and Phylogenomic Analysis of Phototrophic Euglenoids. *Journal of Eukaryotic Microbiology*, *65*(5), 648–660. <https://doi.org/10.1111/jeu.12510>
- Dobáková, E., Flegontov, P., Skalický, T., & Lukeš, J. (2015). Unexpectedly streamlined mitochondrial genome of the euglenozoan *Euglena gracilis*. *Genome Biology and Evolution*, *7*(12), 3358–3367. <https://doi.org/10.1093/gbe/evv229>
- Ebenezer, T. E., Zoltner, M., Burrell, A., Nenarokova, A., Novák Vanclová, A. M. G., Prasad, B., Soukal, P., Santana-Molina, C., O'Neill, E., Nankissoor, N. N., Vadakedath, N., Daiker, V., Obado, S., Silva-Pereira, S., Jackson, A. P., Devos, D. P., Lukeš, J., Lebert, M., Vaughan, S., ... Field, M. C. (2019). Transcriptome, proteome and draft genome of *Euglena gracilis*. *BMC Biology*, *17*(1), 1–23. <https://doi.org/10.1186/s12915-019-0626-8>
- Gockel, G., Hachtel, W., & Melkonian, M. (2000). Complete gene map of the plastid genome of the nonphotosynthetic Euglenoid flagellate *astasia longa*. *Protist*, *151*(4), 347–351. [https://doi.org/10.1078/S1434-4610\(04\)70033-4](https://doi.org/10.1078/S1434-4610(04)70033-4)
- Hallick, R. B., Hong, L., Drager, R. G., Favreau, M. R., Monfort, A., Orsat, B., Spielmann, A., & Stutz, E. (1993). Complete sequence of *Euglena gracilis* chloroplast DNA. *Nucleic Acids Research*, *21*(15), 3537–3544. <https://doi.org/10.1093/nar/21.15.3537>

- Hammond, M. J., Nenarokova, A., Butenko, A., Dobáková, E. L., Field, M. C., Lukeš, J., Zoltner, M., Dobáková, E. L., Field, M. C., & Lukeš, J. (2020). A Uniquely Complex Mitochondrial Proteome from *Euglena gracilis*. *Molecular Biology and Evolution*, 1–19. <https://doi.org/10.1093/molbev/msaa061>
- Hannaert, V., Saavedra, E., Duffieux, F., Szikora, J. P., Rigden, D. J., Michels, P. A. M., & Opperdoes, F. R. (2003). Plant-like traits associated with metabolism of Trypanosoma parasites. *Proceedings of the National Academy of Sciences of the United States of America*, 100(3), 1067–1071. <https://doi.org/10.1073/pnas.0335769100>
- Hrdá, Š., Fousek, J., Szabová, J., Hampl, V., & Vlček, Č. (2012). The plastid genome of eutreptiella provides a window into the process of secondary endosymbiosis of plastid in euglenids. *PLoS ONE*, 7(3):e33746. <https://doi.org/10.1371/journal.pone.0033746>
- Karnkowska, A., Bennett, M. S., & Triemer, R. E. (2018). Dynamic evolution of inverted repeats in Euglenophyta plastid genomes. *Scientific Reports*, 8(1), 16071. <https://doi.org/10.1038/s41598-018-34457-w>
- Kořený, L., & Oborník, M. (2011). Sequence evidence for the presence of two tetrapyrrole pathways in *Euglena gracilis*. *Genome Biology and Evolution*, 3(1), 359–364. <https://doi.org/10.1093/gbe/evr029>
- Lax, G., Lee, W. J., Eglit, Y., & Simpson, A. (2019). Ploetids Represent Much of the Phylogenetic Diversity of Euglenids. *Protist*, 170(2), 233–257. <https://doi.org/10.1016/j.protis.2019.03.001>
- Lax, G., & Simpson, A. G. B. (2013). Combining Molecular Data with Classical Morphology for Uncultured Phagotrophic Euglenids (Excavata): A Single-Cell Approach. *Journal of Eukaryotic Microbiology*, 60(6), 615–625. <https://doi.org/10.1111/jeu.12068>
- Leander, B. S., Lax, G., Karnkowska, A., & Simpson, A. G. B. (2017). Handbook of the Protists. In *Handbook of the Protists*. <https://doi.org/10.1007/978-3-319-32669-6>
- Leander, B. S., Triemer, R. E., & Farmer, M. A. (2001). Character evolution in heterotrophic euglenids. *European Journal of Protistology*, 37(3), 337–356. <https://doi.org/10.1078/0932-4739-00842>
- Leliaert, F., Smith, D. R., Moreau, H., Herron, M. D., Verbruggen, H., Delwiche, C. F., & De Clerck, O. (2012). Phylogeny and Molecular Evolution of the Green Algae. *Critical Reviews in Plant Sciences*, 31(1), 1–46. <https://doi.org/10.1080/07352689.2011.615705>
- O'Neill, E. C., Trick, M., Hill, L., Rejzek, M., Dusi, R. G., Hamilton, C. J., Zimba, P. V., Henrissat, B., & Field, R. A. (2015). The transcriptome of *Euglena gracilis* reveals unexpected metabolic capabilities for carbohydrate and natural product biochemistry. *Molecular BioSystems*, 11(10), 2808–2820. <https://doi.org/10.1039/c5mb00319a>
- Pombert, J. F., James, E. R., Janouškovec, J., & Keeling, P. J. (2012). Evidence for Transitional Stages in the Evolution of Euglenid Group II Introns and Twintrons in the Monomorpha aenigmatica Plastid Genome. *PLoS ONE*, 7(12), 338–349. <https://doi.org/10.1371/journal.pone.0053433>

- Turmel, M., Gagnon, M. C., O'Kelly, C. J., Otis, C., Lemieux, C., O'Kelly, C. J., Otis, C., Lemieux, C., O'Kelly, C. J., Otis, C., & Lemieux, C. (2009). The chloroplast genomes of the green algae *Pyramimonas*, *Monomastix*, and *Pycnococcus* shed new light on the evolutionary history of prasinophytes and the origin of the secondary chloroplasts of euglenids. *Molecular Biology and Evolution*, 26(3), 631–648. <https://doi.org/10.1093/molbev/msn285>
- Wiegert, K. E., Bennett, M. S., & Triemer, R. E. (2012). Evolution of the Chloroplast Genome in Photosynthetic Euglenoids: A Comparison of *Eutreptia viridis* and *Euglena gracilis* (Euglenophyta). *Protist*, 163(6), 832–843. <https://doi.org/10.1016/j.protis.2012.01.002>
- Yoshida, Y., Tomiyama, T., Maruta, T., Tomita, M., Ishikawa, T., & Arakawa, K. (2016). De novo assembly and comparative transcriptome analysis of *Euglena gracilis* in response to anaerobic conditions. *BMC Genomics*, 17(1), 1–10. <https://doi.org/10.1186/s12864-016-2540-6>
- Záhonová, K., Füssy, Z., Birčák, E., Novák Vanclová, A. M. G., Klimeš, V., Vesteg, M., Krajčovič, J., Oborník, M., & Eliáš, M. (2018). Peculiar features of the plastids of the colourless alga *Euglena longa* and photosynthetic euglenophytes unveiled by transcriptome analyses. *Scientific Reports*, 8(1), 1–15. <https://doi.org/10.1038/s41598-018-35389-1>

7 Curriculum vitae

RNDr. Štěpánka Hrdá

Born Štěpánka Zitková, 21. 11. 1972 in Liberec

Researcher, Ph.D. student

Charles University, Prague, Průmyslová 595, 252 42, Vestec

hrda@natur.cuni.cz

Education:

2011 – 2020: Ph.D. study of Parasitology, Faculty of Science, Charles University, Prague.
Thesis: Evolution of nuclear and plastid genomes in euglenids

2001: *Examen rigorozum*: RNDr. degree awarded

1996 – 1998: Ph.D. study of Parasitology, Faculty of Science, Charles University, Prague.
Thesis: Charakterizace vnitrodruhového a mezidruhového polymorfismu kokcií
Study not completed due to motherhood.

1991 – 1996: MSc study of Parasitology, Faculty of Science, Charles University, Prague.
Thesis: Změny v chování myši způsobené parazitickým prvokem *Toxoplasma gondii*, Mgr.
degree awarded

Employment history:

2007 – 2020: researcher, head of OMICS/Genomic Laboratory

Grants:

GAUK604313 - Hledání genetických stop po sekundární endosymbióze plastidu u osmotrofického euglenida *Rhabdomonas* sp.

8 List of publications

Flegr, J., Hrdá, Š., & Tachezy, J. (1998). The role of psychological factors in questionnaire-based studies on routes of human toxoplasmosis transmission. *Central European Journal of Public Health*, 6(1), 45–50.

Flegr, J., Zitková, Š., Kodym, P., & Frynta, D. (1996). Induction of changes in human behaviour by the parasitic protozoan *Toxoplasma gondii*. *Parasitology*, 113 (1), 49-54. <https://doi.org/10.1017/s0031182000066269>

Flegr, Jaroslav, Hrdá, Š., & Kodym, P. (2005). Influence of latent “asymptomatic” toxoplasmosis on body weight of pregnant women. *Folia Parasitologica*, 52(3), 199–204. doi: 10.14411/fp.2005.026

Hrdá, Š., Fousek, J., Szabová, J., Hampl, V., & Vlček, Č. (2012). The plastid genome of *Eutreptiella* provides a window into the process of secondary endosymbiosis of plastid in euglenids. *PLoS ONE*, 7(3):e33746. <https://doi.org/10.1371/journal.pone.0033746> 7(3):e33746.

Hrdá, Š., Hroudová, M., Vlček, Č., & Hampl, V. (2017). Mitochondrial Genome of Prasinophyte Alga *Pyramimonas parkeae*. *Journal of Eukaryotic Microbiology*, 64(3), 360-369. <https://doi.org/10.1111/jeu.12371>

Hrdá, S., Votýpka, J., Kodym, P., & Flegr, J. (2000). Transient nature of *Toxoplasma gondii*-induced behavioral changes in mice. *The Journal of Parasitology*, 86(4), 657–663. [https://doi: 10.1645/0022-3395\(2000\)086\[0657:TNOTGI\]2.0.CO;2](https://doi: 10.1645/0022-3395(2000)086[0657:TNOTGI]2.0.CO;2)

Kodym, P., Blažek, K., Malý, M., & Hrdá, Š. (2002). Pathogenesis of experimental toxoplasmosis in mice with strains differing in virulence. *Acta Parasitologica*, 47(3), 239-248.

Kodym, P., Hrdá, Š., Machala, L., Rozsypal, H., Staňková, M., & Malý, M. (2006). Prevalence and incidence of toxoplasma infection in HIV-positive patients in the Czech Republic. *Journal of Eukaryotic Microbiology*, 53 Suppl 1, S160-1. <https://doi.org/10.1111/j.1550-7408.2006.00215.x>

Machala, L., Malý, M., Hrdá, Š., Rozsypal, H., Staňková, M., & Kodym, P. (2009). Antibody response of HIV-infected patients to latent, cerebral and recently acquired toxoplasmosis. *European Journal of Clinical Microbiology and Infectious Diseases*, 28(2), 179–182. <https://doi.org/10.1007/s10096-008-0600-9>

Soukal, P., Hrdá, Š., Karnkowska, A., Milanowski, R., Szabová, J., Hroudová, M., Strnad, H., Vlček, Č., Čepička, I., & Hamp, V. (2020). Genome and transcriptome draft of the heterotrophic euglenid *Rhabdomonas costata*. Rukopis v přípravě.

Vanclová, A. M. G., Hadariová, L., Hrdá, Š., & Hampl, V. (2017). Secondary Plastids of Euglenophytes. *Advances in Botanical Research*, 84, 321–358. <https://doi.org/10.1016/bs.abr.2017.06.008>

Záhonová, K., Petrželková, R., Valach, M., Yazaki, E., Tikhonenkov, D. V., Butenko, A., Janouškovec, J., Hrdá, Š., Klimeš, V., Burger, G., Inagaki, Y., Keeling, P. J., Hampl, V., Flegontov, P., Yurchenko, V., & Eliáš, M. (2018). Extensive molecular tinkering in the evolution of the membrane attachment mode of the Rheb GTPase. *Scientific Reports*, 8(1), 0–11. <https://doi.org/10.1038/s41598-018-23575-0>